



OIPE

RAW SEQUENCE LISTING
 PATENT APPLICATION: US/09/975,139

DATE: 03/08/2002
 TIME: 14:10:21

Input Set : A:\23623-7060 Seq Listing.txt
 Output Set: N:\CRF3\03082002\I975139.raw

```

4 <110> APPLICANT: Genencor International, Inc.
5     Schellenberger, Volker
6     Naki, Donald
7     Morrison, Thomas B.
9 <120> TITLE OF INVENTION: INFORMATION RICH LIBRARIES
12 <130> FILE REFERENCE: 23623-7060
14 <140> CURRENT APPLICATION NUMBER: US 09/975,139
15 <141> CURRENT FILING DATE: 2001-10-10
17 <150> PRIOR APPLICATION NUMBER: US 60/239,476
18 <151> PRIOR FILING DATE: 2000-10-10
20 <160> NUMBER OF SEQ ID NOS: 10
22 <170> SOFTWARE: FastSEQ for Windows Version 4.0
24 <210> SEQ ID NO: 1
25 <211> LENGTH: 269
26 <212> TYPE: PRT
27 <213> ORGANISM: Bacillus lentus
29 <220> FEATURE:
30 <223> OTHER INFORMATION: Savinase - subtilisin protease
32 <400> SEQUENCE: 1
33 Ala Gln Ser Val Pro Trp Gly Ile Ser Arg Val Gln Ala Pro Ala Ala
34 1      5      10      15
35 His Asn Arg Gly Leu Thr Gly Ser Gly Val Lys Val Ala Val Leu Asp
36      20      25      30
37 Thr Gly Ile Ser Thr His Pro Asp Leu Asn Ile Arg Gly Gly Ala Ser
38      35      40      45
39 Phe Val Pro Gly Glu Pro Ser Thr Gln Asp Gly Asn Gly His Gly Thr
40      50      55      60
41 His Val Ala Gly Thr Ile Ala Ala Leu Asn Asn Ser Ile Gly Val Leu
42 65      70      75      80
43 Gly Val Ala Pro Ser Ala Glu Leu Tyr Ala Val Lys Val Leu Gly Ala
44      85      90      95
45 Ser Gly Ser Gly Ser Val Ser Ser Ile Ala Gln Gly Leu Glu Trp Ala
46      100     105     110
47 Gly Asn Asn Gly Met His Val Ala Asn Leu Ser Leu Gly Ser Pro Ser
48      115     120     125
49 Pro Ser Ala Thr Leu Glu Gln Ala Val Asn Ser Ala Thr Ser Arg Gly
50      130     135     140
51 Val Leu Val Val Ala Ala Ser Gly Asn Ser Gly Ala Gly Ser Ile Ser
52 145     150     155     160
53 Tyr Pro Ala Arg Tyr Ala Asn Ala Met Ala Val Gly Ala Thr Asp Gln
54      165     170     175
55 Asn Asn Asn Arg Ala Ser Phe Ser Gln Tyr Gly Ala Gly Leu Asp Ile
56      180     185     190

```

RAW SEQUENCE LISTING

PATENT APPLICATION: US/09/975,139

DATE: 03/08/2002

TIME: 14:10:21

Input Set : A:\23623-7060 Seq Listing.txt

Output Set: N:\CRF3\03082002\I975139.raw

```

57 Val Ala Pro Gly Val Asn Val Gln Ser Thr Tyr Pro Gly Ser Thr Tyr
58      195      200      205
59 Ala Ser Leu Asn Gly Thr Ser Met Ala Thr Pro His Val Ala Gly Ala
60      210      215      220
61 Ala Ala Leu Val Lys Gln Lys Asn Pro Ser Trp Ser Asn Val Gln Ile
62 225      230      235      240
63 Arg Asn His Leu Lys Asn Thr Ala Thr Ser Leu Gly Ser Thr Asn Leu
64      245      250      255
65 Tyr Gly Ser Gly Leu Val Asn Ala Glu Ala Ala Thr Arg
66      260      265

```

69 <210> SEQ ID NO: 2

70 <211> LENGTH: 16

71 <212> TYPE: PRT

72 <213> ORGANISM: Artificial Sequence

74 <220> FEATURE:

75 <223> OTHER INFORMATION: Artificial subtilisin reference protein sequence (Fig. 1)

77 <400> SEQUENCE: 2

```

78 Ser Thr Ser Ile Leu Gly Val Ala Ser Ser Ala Ser Leu Leu Gly Val
79 1      5      10      15

```

82 <210> SEQ ID NO: 3

83 <211> LENGTH: 382

84 <212> TYPE: PRT

85 <213> ORGANISM: Aeromonas sobria

87 <220> FEATURE:

88 <223> OTHER INFORMATION: AmpC protein

90 <400> SEQUENCE: 3

```

91 Met Lys Gln Thr Arg Ala Leu Pro Leu Leu Ala Leu Gly Thr Leu Leu
92 1      5      10      15
93 Leu Ala Pro Leu Ser Leu Ala Ala Pro Val Asp Pro Leu Lys Ala Val
94      20      25      30
95 Val Asp Asp Ala Ile Arg Pro Val Leu Lys Gln His Arg Ile Pro Gly
96      35      40      45
97 Met Ala Val Ala Val Leu Lys Gly Gly Gln Ala His Tyr Phe Asn Tyr
98      50      55      60
99 Gly Leu Ala Asp Val Ala Thr Gly Ala Lys Val Asn Glu Gln Thr Leu
100 65      70      75      80
101 Phe Glu Ile Gly Ser Val Ser Lys Thr Tyr Thr Ala Thr Leu Gly Ala
102      85      90      95
103 Tyr Ala Val Val Lys Gly Gly Phe Lys Leu Asp Asp Gln Val Ser Gly
104      100      105      110
105 His Ala Pro Trp Leu Lys Gly Ser Ala Phe Asp Gly Ile Thr Met Ala
106      115      120      125
107 Glu Leu Ala Thr Tyr Ser Ala Gly Gly Leu Pro Leu Gln Phe Pro Asp
108      130      135      140
109 Glu Val Asp Ser Ser Asp Thr Met Arg Ala Tyr Tyr Arg His Trp Thr
110 145      150      155      160
111 Pro Pro Tyr Gln Ala Gly Thr Gln Arg Gln Tyr Ser Asn Pro Ser Ile
112      165      170      175
113 Gly Leu Phe Gly His Leu Ala Ala Ser Ser Leu Gln Gln Pro Phe Ser

```

RAW SEQUENCE LISTING

PATENT APPLICATION: US/09/975,139

DATE: 03/08/2002

TIME: 14:10:21

Input Set : A:\23623-7060 Seq Listing.txt

Output Set: N:\CRF3\03082002\I975139.raw

```

114          180          185          190
115 Thr Leu Met Glu Gln Thr Leu Leu Pro Ala Leu Gly Leu Glu His Thr
116          195          200          205
117 Tyr Leu Gln Val Pro Glu Ala Ala Met Ala Arg Tyr Ala Phe Gly Tyr
118          210          215          220
119 Ser Lys Glu Asp Lys Pro Ile Arg Val Asn Pro Gly Met Leu Ala Asp
120 225          230          235          240
121 Glu Ala Tyr Gly Ile Lys Thr Gly Ser Ala Asp Leu Leu Ala Phe Val
122          245          250          255
123 Lys Ala Asn Ile Ser Gly Val Asp Asp Lys Ala Leu Gln Gln Ala Ile
124          260          265          270
125 Ala Leu Thr His Thr Gly Phe Tyr Arg Ile Gly Glu Met Ser Gln Gly
126          275          280          285
127 Leu Gly Trp Glu Ser Tyr Ala Tyr Pro Val Ser Glu Gln Thr Leu Leu
128          290          295          300
129 Ala Gly Asn Ser Pro Ala Val Ser Leu Lys Ala Asn Pro Val Thr Lys
130 305          310          315          320
131 Phe Glu Thr Pro Ala Ala Pro Gly Ala Met Arg Leu Tyr Asn Lys Thr
132          325          330          335
133 Gly Ser Thr Gly Gly Phe Gly Ala Tyr Val Ala Phe Val Pro Ala Lys
134          340          345          350
135 Gly Ile Gly Ile Val Met Leu Ala Asn Arg Asn Tyr Pro Ile Glu Ala
136          355          360          365
137 Arg Val Ser Ala Ala His Ala Ile Leu Ser Gln Leu Ala Pro
138          370          375          380
141 <210> SEQ ID NO: 4
142 <211> LENGTH: 381
143 <212> TYPE: PRT
144 <213> ORGANISM: Enterobacter cloacae
146 <220> FEATURE:
147 <223> OTHER INFORMATION: AmpC protein
149 <400> SEQUENCE: 4
150 Met Met Arg Lys Ser Leu Cys Cys Ala Leu Leu Leu Gly Ile Ser Cys
151 1          5          10          15
152 Ser Ala Leu Ala Thr Pro Val Ser Glu Lys Gln Leu Ala Glu Val Val
153          20          25          30
154 Ala Asn Thr Ile Thr Pro Leu Met Lys Ala Gln Ser Val Pro Gly Met
155          35          40          45
156 Ala Val Ala Val Ile Tyr Gln Gly Lys Pro His Tyr Tyr Thr Phe Gly
157          50          55          60
158 Lys Ala Asp Ile Ala Ala Asn Lys Pro Val Thr Pro Gln Thr Leu Phe
159 65          70          75          80
160 Glu Leu Gly Ser Ile Ser Lys Thr Phe Thr Gly Val Leu Gly Gly Asp
161          85          90          95
162 Ala Ile Ala Arg Gly Glu Ile Ser Leu Asp Asp Ala Val Thr Arg Tyr
163          100          105          110
164 Trp Pro Gln Leu Thr Gly Lys Gln Trp Gln Gly Ile Arg Met Leu Asp
165          115          120          125
166 Leu Ala Thr Tyr Thr Ala Gly Gly Leu Pro Leu Gln Val Pro Asp Glu

```

RAW SEQUENCE LISTING

PATENT APPLICATION: US/09/975,139

DATE: 03/08/2002

TIME: 14:10:21

Input Set : A:\23623-7060 Seq Listing.txt

Output Set: N:\CRF3\03082002\I975139.raw

```

167      130      135      140
168 Val Thr Asp Asn Ala Ser Leu Leu Arg Phe Tyr Gln Asn Trp Gln Pro
169 145      150      155      160
170 Gln Trp Lys Pro Gly Thr Thr Arg Leu Tyr Ala Asn Ala Ser Ile Gly
171      165      170      175
172 Leu Phe Gly Ala Leu Ala Val Lys Pro Ser Gly Met Pro Tyr Glu Gln
173      180      185      190
174 Ala Met Thr Thr Arg Val Leu Lys Pro Leu Lys Leu Asp His Thr Trp
175      195      200      205
176 Ile Asn Val Pro Lys Ala Glu Glu Ala His Tyr Ala Trp Gly Tyr Arg
177      210      215      220
178 Asp Gly Lys Ala Val Arg Val Ser Pro Gly Met Leu Asp Ala Gln Ala
179 225      230      235      240
180 Tyr Gly Val Lys Thr Asn Val Gln Asp Met Ala Asn Trp Val Met Ala
181      245      250      255
182 Asn Met Ala Pro Glu Asn Val Ala Asp Ala Ser Leu Lys Gln Gly Ile
183      260      265      270
184 Ala Leu Ala Gln Ser Arg Tyr Trp Arg Ile Gly Ser Met Tyr Gln Gly
185      275      280      285
186 Leu Gly Trp Glu Met Leu Asn Trp Pro Val Glu Ala Asn Thr Val Val
187      290      295      300
188 Glu Gly Ser Asp Ser Lys Val Ala Leu Ala Pro Leu Pro Val Ala Glu
189 305      310      315      320
190 Val Asn Pro Pro Ala Pro Pro Val Lys Ala Ser Trp Val His Lys Thr
191      325      330      335
192 Gly Ser Thr Gly Gly Phe Gly Ser Tyr Val Ala Phe Ile Pro Glu Lys
193      340      345      350
194 Gln Ile Gly Ile Val Met Leu Ala Asn Thr Ser Tyr Pro Asn Pro Ala
195      355      360      365
196 Arg Val Glu Ala Ala Tyr His Ile Leu Glu Ala Leu Gln
197      370      375      380
200 <210> SEQ ID NO: 5
201 <211> LENGTH: 381
202 <212> TYPE: PRT
203 <213> ORGANISM: Escherichia coli
205 <220> FEATURE:
206 <223> OTHER INFORMATION: AmpC protein
208 <400> SEQUENCE: 5
209 Met Met Lys Lys Ser Leu Cys Cys Ala Leu Leu Leu Thr Ala Ser Phe
210 1      5      10      15
211 Ser Thr Phe Ala Ala Lys Thr Glu Gln Gln Ile Ala Asp Ile Val
212      20      25      30
213 Asn Arg Thr Ile Thr Pro Leu Met Gln Glu Gln Ala Ile Pro Gly Met
214      35      40      45
215 Ala Val Ala Val Ile Tyr Gln Gly Lys Pro Tyr Tyr Phe Thr Trp Gly
216      50      55      60
217 Lys Ala Asp Ile Ala Asn Asn His Pro Val Thr Gln Gln Thr Leu Phe
218 65      70      75      80
219 Glu Leu Gly Ser Val Ser Lys Thr Phe Asn Gly Val Leu Gly Gly Asp

```

RAW SEQUENCE LISTING

PATENT APPLICATION: US/09/975,139

DATE: 03/08/2002

TIME: 14:10:21

Input Set : A:\23623-7060 Seq Listing.txt

Output Set: N:\CRF3\03082002\I975139.raw

```

220          85          90          95
221 Ala Ile Ala Arg Gly Glu Ile Lys Leu Ser Asp Pro Val Thr Lys Tyr
222          100          105          110
223 Trp Pro Glu Leu Thr Gly Lys Gln Trp Gln Gly Ile Arg Leu Leu His
224          115          120          125
225 Leu Ala Thr Tyr Thr Ala Gly Gly Leu Pro Leu Gln Ile Pro Asp Asp
226          130          135          140
227 Val Arg Asp Lys Ala Ala Leu Leu His Phe Tyr Gln Asn Trp Gln Pro
228 145          150          155          160
229 Gln Trp Thr Pro Gly Ala Lys Arg Leu Tyr Ala Asn Ser Ser Ile Gly
230          165          170          175
231 Leu Phe Gly Glu Leu Ala Val Lys Pro Ser Gly Met Ser Tyr Glu Glu
232          180          185          190
233 Ala Met Thr Arg Arg Val Leu Gln Pro Leu Lys Leu Ala His Thr Trp
234          195          200          205
235 Ile Thr Val Pro Gln Asn Glu Gln Lys Asp Tyr Ala Trp Gly Tyr Arg
236          210          215          220
237 Glu Gly Lys Pro Val His Val Ser Pro Gly Gln Leu Asp Ala Glu Ala
238 225          230          235          240
239 Tyr Gly Val Lys Ser Ser Val Ile Asp Met Ala Arg Trp Val Gln Ala
240          245          250          255
241 Asn Met Asp Ala Ser His Val Gln Glu Lys Thr Leu Gln Gln Gly Ile
242          260          265          270
243 Ala Leu Ala Gln Ser Arg Tyr Trp Arg Ile Gly Asp Met Tyr Gln Gly
244          275          280          285
245 Leu Gly Trp Glu Met Leu Asn Trp Pro Leu Lys Ala Asp Ser Ile Ile
246          290          295          300
247 Asn Gly Ser Asp Ser Lys Val Ala Leu Ala Ala Leu Pro Ala Val Glu
248 305          310          315          320
249 Val Asn Pro Pro Ala Pro Ala Val Lys Ala Ser Trp Val His Lys Thr
250          325          330          335
251 Gly Ser Thr Gly Gly Phe Gly Ser Tyr Val Ala Phe Val Pro Glu Lys
252          340          345          350
253 Asn Leu Gly Ile Val Met Leu Ala Asn Lys Ser Tyr Pro Asn Pro Val
254          355          360          365
255 Arg Val Glu Ala Ala Trp Arg Ile Leu Glu Lys Leu Gln
256          370          375          380
259 <210> SEQ ID NO: 6
260 <211> LENGTH: 390
261 <212> TYPE: PRT
262 <213> ORGANISM: Ochrobactrum anthropi
264 <220> FEATURE:
265 <223> OTHER INFORMATION: AmpC protein
267 <400> SEQUENCE: 6
268 Met Arg Thr Ser Thr Thr Leu Leu Ile Gly Phe Leu Thr Thr Ala Ala
269 1          5          10          15
270 Val Ile Pro Asn Asn Gly Ala Leu Ala Ala Ser Lys Val Asn Asp Gly
271          20          25          30
272 Asp Leu Arg Arg Ile Val Asp Glu Thr Val Arg Pro Leu Met Ala Glu

```

Use of n and/or Xaa has been detected in the
Sequence Listing. Review the Sequence Listing
to ensure a corresponding explanation is present
in the <220> to <223> fields of each sequence
using n or Xaa

VERIFICATION SUMMARY

PATENT APPLICATION: US/09/975,139

DATE: 03/08/2002

TIME: 14:10:22

Input Set : A:\23623-7060 Seq Listing.txt

Output Set: N:\CRF3\03082002\I975139.raw

L:524 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:10
L:526 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:10
L:528 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:10
L:530 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:10
L:532 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:10
L:536 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:10
L:538 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:10
L:542 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:10
L:544 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:10
L:546 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:10
L:548 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:10
L:550 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:10
L:552 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:10
L:554 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:10
L:556 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:10
L:558 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:10
L:560 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:10
L:562 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:10
L:564 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:10
L:566 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:10
L:568 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:10
L:570 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:10
L:572 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:10